Dense Disparity Estimation From Stereo Images

Wided Miled, Jean-Christophe Pesquet, Michel Null Parent

DENSE DISPARITY ESTIMATION FROM STEREO IMAGES

Wided Miled [1,2], *Jean Christophe Pesquet* [2] *and Michel Parent* [1]

[1]INRIA, IMARA Project
Domaine de Voluceau
78150 Le chesnay, Cedex France
email : {wided.miled,michel.parent}@inria.fr

[2]Institut Gaspard Monge and UMR-CNRS 8049
Université Marne-la-Vallée, Champs-sur-Marne
77454 Marne-la-Vallée, France
e-mail: pesquet@univ-mlv.fr

ABSTRACT

Binocular stereovision is based on the process of obtaining the depth information from a pair of left and right views of a scene. In this paper, we describe a new approach for computing a dense disparity field based on a convex set theoretic formulation. The stereo matching problem is solved through the minimization of an objective function under various convex constraints arising from prior knowledge. In order to preserve the discontinuities in the disparity field while getting a stable solution, we consider different appropriate regularization constraints. The resulting multi-constrained optimization problem is solved via an efficient algorithm which was recently introduced in the convex optimization literature. Experimental results on standard data sets demonstrate the validity of the proposed approach.

1. INTRODUCTION

One of the classical research problems in the field of computer vision is that of stereo, i.e, reconstruction of the 3D geometry of a scene from a pair of left and right 2D views. Stereo vision has a wide range of applications, including video coding, view synthesis, object recognition and safe navigation in spatial environments. The main goal of stereo vision is to find corresponding pixels, i.e. pixels resulting from the projection of the same 3D point onto the two image planes. The difference in location between corresponding pixels forms the so-called disparity map.

There are several factors that make disparity estimation a difficult task, including depth discontinuities, photometric variation and lack of texture. The presence of discontinuities causes occlusions, which are points only visible in one image of the stereo pair, making the disparity assignment very difficult at object boundaries. Furthermore, image areas which contain little or repetitive textures result in ambiguities in the matching process caused by the presence of multiple possible matches.

In order to overcome these ambiguities and make the problem more tractable, a variety of constraints and assumptions are typically made. The most commonly used constraints are related to the following factors:

- *Epipolar geometry*: reduces the stereo correspondence problem to a one-dimensional one, by involving that matched points always occur on homologous epipolar lines. This problem can be further simplified by rectifying the images so that epipolar lines lie parallel to the image rasters [1].
- *Intensity constancy*: states that a corresponding pixel pair should have the same image intensity value.
- *Uniqueness*: imposes that one pixel can have, at most, one corresponding pixel in the other image.
- *Ordering*: constrains the order of points along epipolar lines to remain the same.
- *Smoothness*: imposes a continuous and smooth variation in the uniform areas of the disparity field.

A broad range of algorithms, using combinations of the above constraints, have been developed for solving the stereo matching problem. Stereo algorithms that provide sparse disparity maps are based on the matching of extracted salient features from both images, such as edges, corners or segments. A good survey for different feature matching strategies is addressed in [2]. Those methods have a low computational complexity and are well suited for real time processing tasks. In addition, they allow the recovery of large displacements and establish accurate disparity estimates. However, methods computing only sparse matches cannot be considered in many applications of stereo. The work of [3] combines the reliability of feature-based methods with the resolution of dense approaches. This is also the idea behind the work presented in [4], where matched features guide the subsequent matching process within a progressive framework.

Stereo algorithms that produce dense disparity maps can be further classified as local or global optimization methods. Local algorithms, where the disparity at each pixel depends only on intensity values within a local window, perform well in highly textured regions, however they often produce noisy disparities in textureless regions and fail at occluded areas. Global algorithms make smoothness assumptions and solve the problem through various minimization techniques. These optimization methods manipulate a large amount of data, which implies a high computational cost. In order to circumvent this disadvantage, multiresolution representations based on wavelet analysis are often proposed [5]. Thus, the matching process is performed by a coarse-to-fine scheme, where finer disparity maps are successively refined from coarser ones.

In this paper, we present a new approach for disparity estimation based on a set theoretic formulation. The proposed method is a global stereo method inspired from recent work developed for image restoration purposes [6] [7]. The stereo matching problem is solved through minimizing a quadratic objective function under certain convex constraints. In order to obtain a smooth disparity field while preserving discontinuities, both total variation and wavelet based constraints are considered. The resulting optimization problem is solved with a block iterative decomposition method which offers great flexibility in the incorporation of additional constraints and allows the combination of both spatial and wavelet domain constraints.

The outline of the paper is as follows. Section 2 first provides a quick overview of some dense stereo methods. In Section 3, we introduce our set theoretic disparity estimation approach, mainly focusing on convex regularization constraints. We provide several experimental results in Section 4. Finally, some concluding remarks are given in Section 5.

2. DENSE STEREO METHODS

In this section, we briefly review the main approaches used to provide dense disparity maps, namely local and global methods. For a complete survey of the state-of-the art in computational dense stereo, we refer the reader to [9].

2.1. Local correspondence methods

Although many local methods have been developed, we only present the correlation based method which summarizes the approach used by most algorithms. Let I_l and I_r be the left and right intensity images of a rectified stereo pair. Given a pixel in the left image, the problem is to find the corresponding pixel in the right image by minimizing a cost function over a correlation search window:

$$\widehat{u}(x,y) = \arg\min_{u\in\Omega} \ \tilde{J}\left(x, y, u(x,y)\right) , \tag{1}$$

where $\widehat{u}$ is the estimated disparity field, Ω is the range of candidate disparity vectors and $\tilde{J}$ is a cost function. Many different similarity measures have been used in the literature. The most commonly used are the Sum of Square Differences (SSD), the Sum of Absolute Differences (SAD) and the Normalized Cross-Correlation (NCC).

Usually, the selection of the appropriate size of the search window is not a trivial task. Moreover, the use of windows of fixed size and shape may lead to erroneous matches in the most challenging image regions. The different approaches for adaptive/shiftable windows [10] attempt to solve this problem by varying the size and shape of the window according to the intensity variation. The work in [11] uses a multiple window method where a number of distinct windows are tried and the one providing highest correlation is retained.

2.2. Global optimization methods

The second general approach to deal with ambiguities in stereo correspondence is to optimize a global energy function. Typically, such a function consists of two terms and takes the following form:

$$E(u) = E_{\mathrm{data}}(u) + \lambda\, E_{\mathrm{smooth}}(u). \tag{2}$$

The first term measures the distance between corresponding pixels, while the second one enforces the smoothness of the disparity field and λ is a positive constant weighting the two terms. Several different energy minimization algorithms have been proposed to solve Equation (2). The most common approach is dynamic programming, which uses the ordering and smoothness constraints to optimize correspondences in each scanline. The matching costs of all points on a scanline describe the disparity search space. Finding the correct disparities is akin to finding the minimum cost path through this space. The most significant limitation of dynamic programming is its inability to enforce smoothness in both horizontal and vertical directions. The work of [14] proposes a way to cope with this problem while maintaining the dynamic programming framework. Recently, powerful algorithms have been developed based on graph cuts and belief propagation for minimizing the full 2D global energy function [10]. the idea is to cast the stereo matching problem as a pixel labelling problem to find the minimum cut through a certain graph.

Variational approaches have also been very effective for minimizing Equation (2) via an iterative scheme derived from the associated Euler-Lagrange differential equations [12], [13]. However, these techniques often are computationally demanding and they require a careful study for the discretization of the associated partial differential equations. Besides, they require the determination of the Lagrange parameter λ which may be a difficult task. The latter problem becomes even more involved when a sum of regularization terms has to be considered to address multiple constraints which may arise in the problem.

3. PROPOSED APPROACH

In this section, we describe the formulation of the stereo matching as a constrained optimization problem and we introduce the two basic elements required to solve this problem, namely the objective function and the convex constraints.

3.1. Formulation of the correspondence problem

Let the cost function be the SSD. Finding a corresponding pixel in the right image I_r for each pixel in the left image I_l amounts to search the disparity u that minimizes:

$$\tilde{J}(u) = \sum_{(x,y)\in\mathcal{D}} \left[I_l(x,y) - I_r(x - u(x,y), y)\right]^2 , \tag{3}$$

where $\mathcal{D} \subset \mathbb{N}^2$ is the image support. To find $\widehat{u} = \arg\min_u \ \tilde{J}(u)$, a tough non-convex minimization problem has therefore to be solved. However to circumvent this difficulty, we assume that an initial estimate $\bar{u}$ of u is available, for example from a correlation based method and we compensate the non-linear term $I_r(x-u,y)$ around $\bar{u}$ using the standard first order approximation:

$$I_r(x-u,y) \simeq I_r(x-\bar{u},y) - (u-\bar{u}) \ \ I_r^x(x-\bar{u},y) , \tag{4}$$

where $I_r^x(x-\bar{u},y)$ is the horizontal gradient of the warped right image. Note that for notation concision, we have not made explicit that u and $\bar{u}$ are functions of (x,y) in the above expression.

This linearization leads to the following convex quadratic criterion:

$$J(u) = \sum_{(x,y)\in\mathcal{D}} \left[L(x,y)\; u - r(x,y)\right]^2 , \tag{5}$$

where

$$L(x,y) = I_r^x(x-\bar{u},y), \tag{6}$$
$$r(x,y) = I_r(x-\bar{u},y) + \bar{u}\; L(x,y) - I_l(x,y). \tag{7}$$

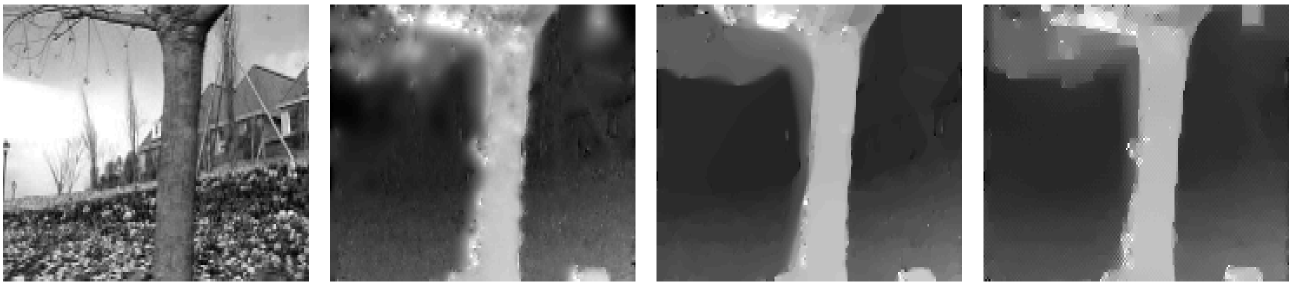
Now, our purpose is to estimate the disparity image u from the observed fields L and r. This problem is often ill-posed and it must therefore be regularized by adding some constraints modelling the available information. In this work, we address this problem from a set theoretic formulation, where each piece of information is represented by a convex set in the solution space and the intersection of these sets, the feasibility set, constitutes the family of possible solutions. The aim is to find an acceptable solution minimizing a certain cost function, the problem being formulated in a Hilbert image space $\mathcal{H}$ as follows:

$$\text{Find } u \in S = \bigcap_{i=1}^{m} S_i \text{ such that } J(u) = \inf J(S) , \tag{8}$$

where the objective $J : \mathcal{H} \to]-\infty, +\infty]$ is a convex function and the constraint sets $(S_i)_{1\le i\le m}$ are closed convex subsets of $\mathcal{H}$. Constraint sets can be represented as level sets:

$$\forall i \in \{1, ..., m\}, \qquad S_i = \{u \in \mathcal{H} \mid f_i(u) \le \delta_i\} , \tag{9}$$

where $f_i : \mathcal{H} \to]-\infty, +\infty]$ is a convex function and $\delta_i \in \mathbb{R}$. Several methods have been proposed to solve the convex feasibility problem in (8). In this work, we use the constrained quadratic



(a) (b) (c) (d)

Figure 1. (a) Reference flower garden image. Extracted disparity maps using (b) Nagel-Enkelmann constraint (c) Total variation constraint (d) wavelet based constraint.

minimization method developed in [8], which is particularly well adapted to our need. Due to space limitation, we will not describe the algorithm but the reader is referred to [8] for more details.

In what follows, we introduce our objective function as well as the considered convex constraints to solve the stereo matching problem within the framework described above.

3.2. Energy function

The objective function is the quadratic measure modelling our linearized data formulation. To avoid large deviations from the data model, we exclude occluded pixels from the summation of pixel matching cost in Equation (5). Note that, although these points are discarded from the objective function, they are taken into account in the expression of the constraints that will be discussed in Section 3.3. Denoting $\mathcal{O}$ the occlusion field and $\mathbf{x} = (x, y)$ a pixel in the left view, the energy function to be minimized is given by:

$$J(u) = \sum_{\mathbf{x}\in\mathcal{D}\setminus\mathcal{O}} [L(\mathbf{x})\, u(\mathbf{x}) - r(\mathbf{x})]^2 + \alpha \sum_{\mathbf{x}\in\mathcal{D}} [u(\mathbf{x}) - \bar{u}(\mathbf{x})]^2 \ , \tag{10}$$

where $\bar{u}$ is an initial estimate and α is a small positive constant that weights the first term relatively to the second. The primary role of this term is not to regularize the solution but to make J strictly convex, in compliance with the assumption required to guarantee the convergence of the algorithm we use [8]. To further address the occlusion problem, we have to deal with a consistent initial disparity field. Based on the correlation based method described in Section 2.1, we first compute left-to-right and right-to-left initial disparity maps, denoted respectively by $\bar{u}_l$ and $\bar{u}_r$, and for each point **x** we take $\bar{u}(\mathbf{x}) = \bar{u}_r(x - \bar{u}_l(\mathbf{x}), y)$. We then iteratively refine the initial disparity estimate by choosing the result from a previous estimate as the initial value of the next step, which further reduces the sensitivity of the final solution to the initial estimate.

3.3. Convex constraints

The construction of convex constraints is derived from the properties of the estimated field. The constraint set arising from the knowledge of the disparity range values is $S_1 = [u_{\min}, u_{\max}]$. We note that in practice $u_{\min}$ and $u_{\max}$ are often available. For the regularization constraint, we experiment spatial and wavelet domain constraints. First, we consider an oriented smoothness constraint inspired from the work in [12]. This constraint based on the Nagel–Enkelmann operator $(\nabla u)^\top D(\nabla I_l)(\nabla u)$, introduces an anisotropic behavior to avoid an oversmoothing along object edges. Details given in [12] provide guidelines for obtaining an approximation κ of this term leading to the following convex set

$$S_2 = \{u \in \mathcal{H} \mid (\nabla u)^\top D(\nabla I_l)(\nabla u) \leq \kappa\} \ . \tag{11}$$

Secondly, we make use of the Total Variation (TV) measure which recently attracted much attention as it achieves good results in various research fields such as image restoration [7]. Controlling the TV amounts to smooth uniformly homogeneous regions while preserving sharp edges. Hence, imposing an upper bound τ on the total variation of u, restricts the solutions to the convex set

$$S_3 = \{u \in \mathcal{H} \mid \mathrm{TV}(u) \leq \tau\} \ . \tag{12}$$

It should be noted that $\mathrm{TV}(u)$ constitutes a geometrical feature that can be estimated from experiments and image databases for a given class of images.

Finally, we adopt a wavelet domain approach to construct a regularization constraint based on 2D separable wavelet coefficients. Inspired from the recent work proposed in [6], we consider the norm of Besov space $B^1_{1,1}$ as it is appropriate to model images that contain discontinuities. The convex set associated to a semi-norm of this space is given by

$$S_4 = \{u \in \mathcal{H} \mid \sum_{j,k,o} |w^{\mathcal{B}}_{j,o,k}(u)| \leq \nu\}, \tag{13}$$

where $(w^{\mathcal{B}}_{j,o,k}(u))_{k\in\mathbb{Z}^2}$ denote the wavelet coefficients of $u \in \mathcal{H}$, $j \in \mathbb{Z}$ is the resolution level and $o \in \{1, 2\}$ is the orientation parameter corresponding to the horizontal and vertical directions in the image plane. Note that the upper bound ν can be estimated from experiments on available databases.

To illustrate the behavior of the above regularization constraints, we consider a pair of images of the real flower garden sequence. Results reported in Figure 1 show that Total Variation and wavelet based constraints clearly outperform the Nagel-Enkelmann based constraint. However, to benefit from the ability of the applied optimization method to process several constraints, we consider two scenarios. First, we minimize J in (10) over $S = \bigcap_{i=1}^{3} S_i$ and secondly we consider the feasibility set to be $S = \bigcap_{i\in\{1,2,4\}} S_i$. This allows to improve the results in both cases. Furthermore, the second scenario is observed to lead to a faster convergence speed than the first one.

4. EXPERIMENTAL RESULTS

In this section, we report the results of our approach using the standard data sets available at the Middlebury website [9]. Table 1 illustrates a quantitative comparison with results from other stereo algorithms available at the same website using the Mean Absolute Error (MAE) measure. Table 1 also includes the results provided by our method using the two scenarios described

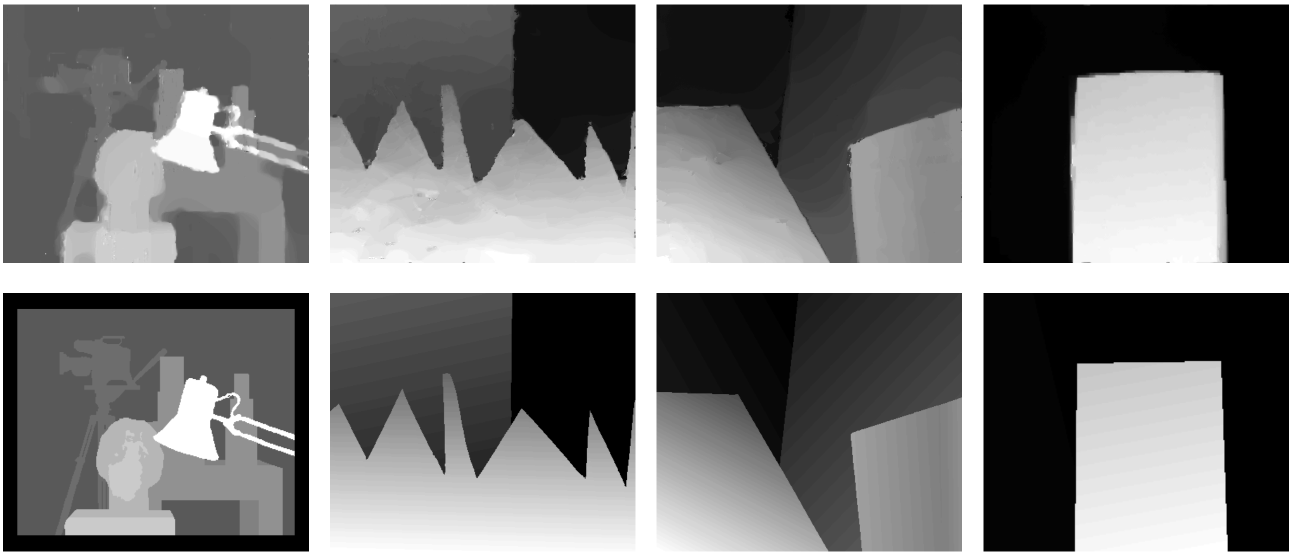


Figure 2. Results on Middlebury data set. From top to bottom: computed disparity maps and ground truth disparity maps. From left to right: Tsukuba, Sawtooth, Venus, Map.

Technique	Tsukuba	Sawtooth	Venus	Map
Segm. +glob.vis. [14]	**0.18**	0.24	0.31	0.67
2-pass DP [15]	0.28	0.25	0.30	0.37
Proposed ($S = \bigcap_{i=1}^{3} S_i$)	0.29	**0.23**	**0.24**	**0.35**
Proposed ($S = \bigcap_{i\in\{1,2,4\}} S_i$)	0.40	0.32	0.44	0.58

Table 1. Performance comparison of stereo algorithms using the MAE measure.

in Section 3.3. Following the evaluation procedure in [9], we only consider non-occluded pixels when computing the disparity error. An inspection of the values of the MAE reveals that the Total Variation constraint yields better results than the wavelet one. It is worth noting that the wavelet based regularization constraint can be much improved using a translation invariant wavelet representation. Furthermore, the comparison in Table 1 indicates that the performance of our Total Variation based method is comparable with the state-of-the-art stereo algorithms [14], [15]. Figure 2 shows the computed dense disparity maps based on TV regularization constraint as well as the ground truths. We obtain good results for Sawtooth, Venus and Map data sets, but less satisfactory ones for the very complex Tsukuba stereo pair.

5. CONCLUSION

In this paper, we presented a new method for estimating a dense disparity from stereo images by directly formulating the problem as a constrained optimization problem. Within the proposed framework, a quadratic objective function is minimized under convex constraints. Various discontinuity-preserving regularization constraints have been considered. Our evaluations show that the proposed approach performs very well compared with existing methods.

6. REFERENCES

[1] Fusiello, A. and Roberto, V. and Trucco, E., " A compact algorithm for rectification of stereo pairs, "*Machine Vision and Applications*, Vol. 12, n. 1, pp 16-22, 2000.

[2] R. Laganière and E. Vincent, "Matching feature points in stereo pairs: A comparative study of some matching strategies,"*Machine Graphics and Vision*, vol. 10, pp. 237–259, 2001.

[3] J. Konrad and Z.D. Lan, "Dense Disparity estimation from feature correspondences, " in *SPIE Stereoscopic Displays and Virtual Reality Syst.*, vol. 3957, pp. 90-101, Jan. 2000.

[4] Y. Wei and L. Quan, "Region-based progressive stereo matching," in *Computer Vision and Pattern Recognition*, vol. 1, pp. 106–113, 2004.

[5] Y. S. Kim, J.J. Lee, and Y.H. Ha, "Stereo matching algorithm based on modified wavelet decomposition Process,"*Pattern Recognition*, vol. 30, pp. 929–952, 1997.

[6] P. L. Combettes and J. C. Pesquet, "Wavelet-constrained image restoration," *Int. Journal of Wavelets, Multiresolution and Information Proc.*, vol. 2, pp. 371–389, Dec. 2004.

[7] P. L. Combettes and J. C. Pesquet, "Image restoration subject to a total variation constraint,"*IEEE Trans. Image Proc.*, Vol. 13, pp. 1213-1222, Sept. 2004.

[8] P.L. Combettes, "A block iterative surrogate constraint splitting method for quadratic signal recovery, "*IEEE Trans. on Signal Proc.*, Vol. 51, pp. 1771-1782, Jul. 1997.

[9] D. Scharstein and R. Szeliski, "A Taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. Journal of Computer vision*, vol. 47, pp. 7–42, 2002.

[10] S.B. Kang, R. Szeliski and J. Chai, "Handling occlusions in dense multi-view stereo," in *Computer Vision and Pattern Recognition*, vol. 1, pp. 156–161, 2002.

[11] A.Fusiello, V. Roberto and E. Trucco, "Symmetric stereo with multiple windowing," *Int. Journal of Pattern Recognition and Artificial Intelligence*, vol. 14, no. 8, pp. 1053–1066, 2000.

[12] L. Alvarez, R. Deriche, J. Sanchez and J. Weickert, 'Dense disparity map estimation respecting image discontinuities: A PDE and scale-space based approach," *Journal of Visual Communication and Image Representation*, Vol. 13, pp. 3–21, 2002.

[13] Aubert, G. and Deriche, R. and Kornprobst, P., "Computing Optical Flow via Variational Techniques," *SIAM Journal on Numerical Analysis*, pp 156-182, 1999.

[14] C. Kim, K.M. Lee, B.T. Choi, and S.U. Lee, "A dense stereo matching using two-pass dynamic programming with generalized ground control points,"*Computer Vision and Pattern Recognition*, vol. 2, pp. 1075-1082, 2005.

[15] M. Bleyer and M. Gelautz, "A layered stereo algorithm using image segmentation and global visibility constraints,"*Int. Conf. on Image Proc.*, pp. 2997-3000, 2004.